

AMENDMENT TO THE CLAIMS

1.(Currently Amended) A method for identifying at least two component words in a compound word, comprising:

receiving the compound word;

searching a lexicon for the compound word, the lexicon including word entries that are annotated with associated segmentation information indicative of whether the associated word can form a component of a compound word; and

identifying the component words for the compound word based on the entries contained in the lexicon including the annotated word entries: by matching a first word portion of the compound word with a first entry in the lexicon, and matching a second word portion of the compound word with a second entry in the lexicon wherein matching the first word portion further comprises:

searching the lexicon for a word entry that matches the first word portion;

if a match is found;

analyzing the second word portion for matches with word entries in the lexicon;

if a match is found for the second word portion;

generating a hypotheses and adding the hypothesis to a list of hypotheses, wherein the hypothesis is a combination of the first word portion and the second word portion;

prior to adding the hypothesis to the list of hypotheses, checking if the matching word entry in the lexicon for the second word portion or a last portion of the second portion includes an anti-seg indication as the segmentation information indicating that the associated word cannot be a final component of a compound word; and

adding the hypothesis to the list of hypotheses only if the entry does not include the anti-seg indication.

2. Canceled.

3. (Original) The method of claim ~~2~~1 further comprising:
matching additional portions of the compound word until reaching a final character of the compound word.

4. (Original) The method of claim 3 further comprising:
generating a first hypothesis of the component words of the compound word, wherein the first hypothesis is a combination of the matched portions of the compound word.

5. (Original) The method of claim 4 further comprising:
generating additional hypotheses of the component words of the compound word.

6. (Original) The method of claim 5 further comprising:
ranking each identified hypothesis based on a relative likelihood of being a correct representation of the component words of the compound word.

7. (Original) The method of claim 6 wherein ranking is based on data obtained through statistical analysis.

8. (Currently Amended) The method of claim 5 wherein matching the first word portion comprises matching characters in the compound word starting from a first character of the compound word; and

wherein matching the second word portion comprises matching characters in the compound word starting from a first character that follows a last character of the first word portion.

9. Canceled.

10. (Currently Amended) The method of claim 9-8 wherein if additional matches are found for the first and second word portions in the lexicon, adding these matches as alternative hypotheses to the list of hypotheses.

11. (Currently Amended) The method of claim 9-8 further comprising:
adding the hypothesis to the list of hypotheses regardless of whether the segmentation information in the matched word entries for the first word portion and the second word portion indicates that associated words in the matched word entries can form a component of a compound word; and
excluding a resulting trace as invalid at a final evaluation of hypotheses if the segmentation information in either matched word entry indicates that the associated word cannot form a component of a compound word.

12. (Currently Amended) The method of claim 11 wherein if the matched word entry for the second word portion includes an associated word which is a final component in the compound word, further comprising the steps of:

checking if the matched word entry for the final component includes the segmentation information indicating that the associated word can be a final component of a compound word;
adding the hypothesis to the list of hypotheses; and
excluding the resulting trace as invalid at the final evaluation of hypotheses, if ~~an~~ the anti-seg bit is present as the segmentation information indicating that the associated word in the word entry matching the final component of the compound word cannot form a final component of the compound word.

13. (Currently Amended) The method of claim 9-8 further comprising:
returning to the first portion;

adding a character following the last character in the first portion to the first portion;
repeating the steps of searching, generating, and analyzing;
if additional matching word entries are found for the first and second portions in the
lexicon, adding associated words in these matching word entries as alternative
hypotheses to the list of hypotheses.

14. (Currently Amended) The method of claim ~~9~~8 further comprising:
prior to adding the hypothesis to the list of hypotheses, checking if the matching word entry
in the lexicon for the first portion includes segment information indicating that the
word associated with the matching word entry can be a component of a compound
word; and
adding the hypothesis to the list of hypotheses only if the entry includes the segmentation
information.

15. (Previously Amended) The method of claim 14 wherein if the second portion includes a
portion which is a non-final portion, checking if matching word entries for the non-final portion
includes the segmentation information; and
adding the hypothesis to the list of hypotheses only if the matching word entries for all non-
final segments include the segmentation information.

16. Canceled.

17. (Currently Amended) A tangible computer readable medium containing computer
executable instructions that, when executed, cause a computer to perform the steps of:
receiving the compound word;
searching a lexicon for entries matching the compound word, the entries including words
and associated segmentation indicators indicating whether the associated words can
be a component part of a compound word;

identifying component words for the compound word based on the entries contained in the lexicon by matching a first portion of the compound word with a first entry in the lexicon that can be a component of a compound word, as indicated by the associated segmentation indicator; and

matching a second portion of the compound word with a second entry in the lexicon that can be a component of a compound word, as indicated by the associated segmentation indicator; and

if either the first or second portions of the compound word match words in the entries in the lexicon, ~~word match words in the entries the lexicon,~~ but the associated segmentation indicator in either of the matched entries indicates that the associated word cannot be a component of a compound word, then searching for a variation of the associated word in the lexicon that has a segmentation indicator indicating the variation can be a component of a compound word.

18. (Original) The computer readable medium of claim 17 further comprising instructions to perform the steps of:

matching additional portions of the compound word until reaching a final character of the compound word.

19. (Previously Amended) The computer readable medium of claim 18 further comprising instructions to perform the steps of:

generating a first hypothesis of the component words of the compound word, wherein the first hypothesis is a combination of the words in the matched entries in the lexicon; and

generating additional hypotheses of the component words of the compound word if additional matched entries are found in the lexicon.

20. (Original) The computer readable medium of claim 19 further comprising instructions to

perform the steps of:

ranking each identified hypothesis based on a relative likelihood of being a correct representation of the component words of the compound word.

21. (Original) The computer readable medium of claim 17 further comprising instructions to perform the steps of:

wherein matching the first portion comprises matching characters in the compound word starting from a first character of the compound word; and

wherein matching the second portion comprises matching characters in the compound word starting from a first character that follows a last character of the first portion.

22. (Previously Amended) The computer readable medium of claim 21 wherein matching the first portion further comprises instructions to perform the steps of:

searching the lexicon for an entry that matches the first portion;

if a match is found;

analyzing the second portion for matches with entries in the lexicon;

if a match is found for the second portion;

generating a hypothesis in a list of hypotheses, wherein the hypothesis is a combination of the words in the entries matching the first portion and the second portion.

23. (Currently Amended) The computer readable medium of claim 22 wherein if additional matching entries are found for the first and second portions in the lexicon, adding words associated with these matching entries as alternative hypotheses to the list of hypotheses.

24. (Original) The computer readable medium of claim 23 further comprising instructions to perform the steps of:

excluding a resulting hypothesis as invalid at a final evaluation of hypotheses.

25. Canceled.

26. (Currently Amended) A method of spell-checking a compound word, comprising:
searching a lexicon for the compound word, the lexicon including entries having a word
and an indicator indicating whether the associated word can be a component of a
compound word;
if the compound word is not found in the lexicon;
identifying component words comprising the compound word;
comparing each of the identified component words with entries in the lexicon;
determining if each of the component words is correctly used as a component of a
compound word based on the indicator in matched entries in the lexicon;
and
if at least one of the identified component words is used incorrectly, indicating
that the compound word includes a spelling error; and
generating a spelling suggestion for the compound word based on the component words;
wherein determining if each component word is correctly used comprises:
if the component word is not the last component word in the compound, checking
if the indicator for the entry matching the component word indicates that
the word in the matched entry can be a component of a compound word;
if the component word cannot be a component of a compound word, applying
spelling transition rules to the component word and adding no more than
two additional characters to the component word that are not present in the
component word to obtain a new component word; and
rechecking if the new component word is in the lexicon with the indicator.

27. Canceled.

28. Canceled.

29. (Currently Amended) The method of claim ~~28~~26 wherein if applying speller transition rules results in too many unwanted lexical matches, further comprising:

searching the entries in the lexicon for the component word to identify variations of the component word that include the indicator;

generating new compound words that include the identified variations of the component word; and

presenting those new compound words as the spelling suggestion to the user.

30. (Previously Amended) The method of claim 26 wherein determining if each component word is used correctly comprises:

if the component word is the last component word in the compound, checking if the component word has an anti-seg identifier as the indicator, indicating that it cannot be a last component of a compound word;

if the component word has the anti-segment identifier searching the entries in the lexicon for the component word to identify variations of the component word that do not include the anti-seg identifier; and

if an entry with a variation of the component word does not include the anti-seg identifier, suggesting that variation of the component word as the spelling suggestion for the compound word.

31. Canceled.

32. (Currently Amended) The method of claim ~~28~~26 wherein applying speller transition rules in the compound word includes changing characters comprising the component word.

- 33. Canceled.
- 34. Canceled.
- 35. Canceled.

36.(NEW) A method for identifying at least two component words in a compound word, comprising:

receiving the compound word;

searching a lexicon for the compound word, the lexicon including word entries that are annotated with associated segmentation information indicative of whether the associated word can form a component of a compound word; and

identifying the component words for the compound word based on the entries contained in the lexicon including the annotated word entries by searching the lexicon for a word entry that matches a first word portion of the compound word;

if a match is found, analyzing a second word portion of the compound word for matches with word entries in the lexicon;

if a match is found for the second word portion, adding a hypothesis, which is a combination of the first word portion and the second word, to a list of hypotheses, regardless of whether the segmentation information in the matched word entries for the first word portion and the second word portion indicates that associated words in the matched word entries can form a component of a compound word; and

excluding a resulting hypothesis as invalid at an evaluation of the hypotheses if the segmentation information in either matched word entry indicates that the associated word cannot form a component of a compound word;

wherein if the matched word entry for the second word portion includes an associated word which is a final component in the compound word, then further comprising the step of:

excluding the hypothesis as invalid at the evaluation of hypotheses, if an anti-seg bit

is present as the segmentation information indicating that the associated word in the word entry matching the final component of the compound word cannot form a final component of the compound word.

37. (NEW) A method of spell-checking a compound word, comprising:

searching a lexicon for the compound word, the lexicon including entries having a word and an indicator indicating whether the associated word can be a component of a compound word;

if the compound word is not found in the lexicon:

identifying component words comprising the compound word;

comparing each of the identified component words with entries in the lexicon;

determining if each of the component words is correctly used as a component of a compound word based on the indicator in matched entries in the lexicon;

if at least one of the identified component words is used incorrectly, indicating that the compound word includes a spelling error;

wherein determining if each component word is correctly used comprises:

if the component word is not the last component word in the compound word, checking if the indicator for the entry matching the component word indicates that the word in the matched entry can be a component of a compound word;

if the component word cannot be a component of a compound word, applying spelling transition rules to the component word to obtain one or more new component words; and

rechecking if the new component words match words in the lexicon with the indicator;

wherein if applying speller transition rules results in too many unwanted new component words, then presenting new compound words generated from the new component words as the spelling suggestion to the user only when applying speller transition

rules adds, removes, or changes no more than two characters of the component word.

38. (NEW) A method of spell-checking a compound word, comprising:

searching a lexicon for the compound word, the lexicon including entries having a word and an indicator indicating whether the associated word can be a component of a compound word;

if the compound word is not found in the lexicon:

identifying component words comprising the compound word;

comparing each of the identified component words with entries in the lexicon;

determining if each of the component words is correctly used as a component of a compound word based on the indicator in matched entries in the lexicon;

if at least one of the identified component words is used incorrectly, indicating that the compound word includes a spelling error;

wherein determining if each component word is used correctly comprises:

if the component word is not the last component word in the compound word, checking if the indicator for the entry matching the component word indicates that the word in the matched entry can be a component of a compound word;

if the component word cannot be a component of a compound word, applying speller transition rules to the component word to obtain one or more new component words;

rechecking if the new component words are in the lexicon with the indicator; and

generating a spelling suggestion for the compound word based on the new component words;

wherein if applying speller transition rules results in too many unwanted new component words, then further comprising:

searching the entries in the lexicon for the component word to identify variations of the
component word that include the indicator;
generating new compound words that include the identified variations of the component
word; and
presenting those new compound words as the spelling suggestion to the user only if no
more than two component words in the compound word require correction.